

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
26 May 2005 (26.05.2005)

PCT

(10) International Publication Number
WO 2005/048240 A1

(51) International Patent Classification⁷: **G10L 15/18,**
G06F 17/27

(21) International Application Number:
PCT/IB2004/052352

(22) International Filing Date:
9 November 2004 (09.11.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
03104170.0 12 November 2003 (12.11.2003) EP

(71) Applicant (for DE only): **PHILIPS INTELLECTUAL
PROPERTY & STANDARDS GMBH** [DE/DE]; Stein-
damm 94, 20099 Hamburg (DE).

(71) Applicant (for AE, AG, AL, AM, AT, AU, AZ, BA, BB, BE,
BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CY, CZ,

DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD,
SE, SG, SK, SL, SY, SZ, TJ, TM, TN, TR, TT, TZ, UA, UG,
UZ, VC, VN, YU, ZA, ZM, ZW only): **KONINKLIJKE
PHILIPS ELECTRONICS N. V.** [NL/NL]; Groenewoud-
seweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor; and

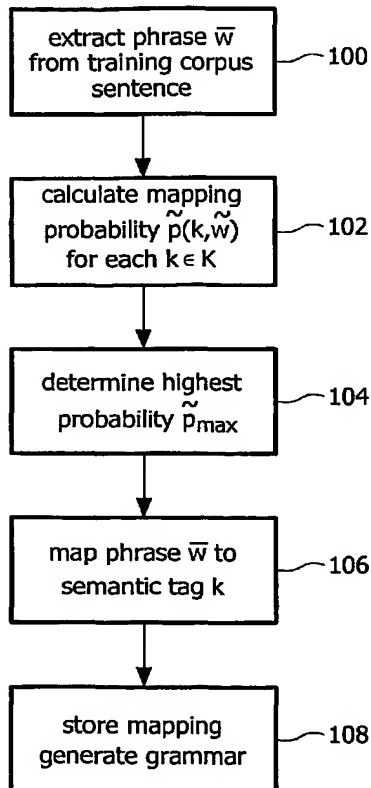
(75) Inventor/Applicant (for US only): **MARTIN, Sven, C.**
[DE/DE]; c/o Philips Intellectual Property &, Standards
GmbH Weissshausstr. 2, 52066 Aachen (DE).

(74) Agents: **MEYER, Michael** et al.; Philips Intellectual
Property &, Standards GmbH Weissshausstr. 2, 52066
Aachen (DE).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,

[Continued on next page]

(54) Title: ASSIGNMENT OF SEMANTIC TAGS TO PHRASES FOR GRAMMAR GENERATION



(57) Abstract: The present invention relates to a method, a system and a computer program product for mapping of semantic tags to phrases within a training corpus of weakly annotated sentences, thereby generating a grammar which can be applied to unknown sentences for the purpose of language understanding. The method is based on a probabilistic estimation that a given phrase is mapped to a semantic tag of a set of candidate semantic tags. The mapping and the generation of the grammar is performed according to a maximum mapping probability of a set of mapping probabilities of the given phrase and the set of candidate semantic tags. In particular, the determination of the mapping probability makes use of an expectation maximization algorithm.

WO 2005/048240 A1



CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.